AFRL-RI-RS-TR-2014-035

CAPTURING COGNITIVE PROCESSING TIME FOR ACTIVE AUTHENTICATION

IOWA STATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

FEBRUARY 2014

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2014-035 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/ S /
ANNA WEEKS
Work Unit Manager

/ S /
WARREN H. DEBANY, JR.
Technical Advisor, Information
Exploitation and Operations Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
FEBRUARY 2014	FINAL TECHNICAL REPORT	MAY 2012 – MAY 2013

4. TITLE AND SUBTITLE

CAPTURING COGNITIVE PROCESSING TIME FOR ACTIVE AUTHENTICATION

5a. CONTRACT NUMBER
FA8750-12-2-0200

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
62788F

6. AUTHOR(S)

Jien Chang

5d. PROJECT NUMBER
ATAU

5e. TASK NUMBER
IO

5f. WORK UNIT NUMBER
WA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Iowa State University of Science and Technology
1350 Beardshear Hall
Ames IA 50011-2025

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RIGA
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER
AFRL-RI-RS-TR-2014-035

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report presents an authentication system that applies machine learning techniques to observe a user's cognitive typing rhythm. A new feature called cognitive typing rhythm (CTR) is used to continuously verify the identities of computer users. Two machine techniques, SVM and KRR, have been developed for the system. The best results from experiments conducted with 1,977 users show a false-rejection rate of 0.7 percent and a false-acceptance rate of 5.5 percent. CTR therefore constitutes a cognitive fingerprint for continuous authentication. Its effectiveness has been verified through a campus-wide experiment at Iowa State University. Furthermore, a live demo was performed twice to demonstrate the effectiveness of our system.

15. SUBJECT TERMS
Active Authentication, Cognitive typing rhythm (CTR), Support Vector Machine (SVM), Behavioral Biometrics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			**ANNA WEEKS**
U	U	U	UU	16	19b. TELEPHONE NUMBER (*Include area code*) **315-330-3936**

TABLE OF CONTENTS

Sections **Pages**

LIST OF FIGURES

Figures **Pages**

1.0 SUMMARY

Conventional authentication systems verify a user only during initial login. Active authentication performs verification continuously as long as the session remains active. This work focuses on using behavioral biometrics, extracted from keystroke dynamics, as "something a user is" for active authentication. This scheme performs continual verification in the background, requires no additional hardware devices, and is invisible to users.

This report presents an authentication system that applies machine learning techniques to observe a user's cognitive typing rhythm. A new feature called cognitive typing rhythm (CTR) is used to continuously verify the identities of computer users. Two machine techniques, SVM and KRR, have been developed for the system. The best results from experiments conducted with 1,977 users show a false-rejection rate of 0.7 percent and a false-acceptance rate of 5.5 percent. CTR therefore constitutes a cognitive fingerprint for continuous authentication. Its effectiveness has been verified through a large-scale dataset.

2.0 INTRODUCTION

Keystroke dynamics—the detailed timing information of keystrokes when using a keyboard—has been studied for the past three decades. The typical keystroke interval time, referred to as a *digraph*, is expressed as the time between typing two characters. A user's keystroke rhythms are distinct enough from person to person for use as biometrics to identify people. However, keystroke rhythm has generally been considered less reliable than physical biometrics, such as fingerprints. The main challenge is the presence of within-user variability.

Owing to this within-user variability of interval times among identical keystrokes, most research efforts have focused on verification techniques that can manage such variability. For example, researchers proposed a method called *degree of disorder* to cope with time variation issues [1,2], arguing that although the keystroke typing durations usually vary between each digraph, the order of the timing tends to be consistent. This suggested that the distance of the order between two keystroke patterns can be used to measure the similarity.

A recent survey on biometric authentication using keystroke dynamics classified research papers on the basis of their feature-extraction methods, feature-subset-selection methods, and classification methods [3]. Most of the systems described in the survey were based on typing rhythms for short sample texts, which are dominated by users' physical characteristics (such as how fast your fingers can move) and are too brief to capture a "cognitive fingerprint." In the current keystroke-authentication commercial market, some products combine the timing information of the password with password-based access control to generate a hardened password [4].

Here, we present a biometric-based active authentication system that continuously monitors and analyzes various keyboard behaviors performed by the user. We extract the features from keystroke dynamics that contain cognitive factors, resulting in cognitive fingerprints. Each feature is a sequence of digraphs from a specific word. This method is driven by our hypothesis that a cognitive factor can affect the typing rhythm of a specific word. Cognitive factors have been largely ignored in previous keystroke dynamics studies.

3.0 METHODS, ASSUMPTIONS AND PROCEDURES

3.1 Cognitive Fingerprint Description

Physical biometrics relies on physical characteristics, such as fingerprints or retinal patterns. The behavioral biometric of keystroke dynamics must incorporate cognitive fingerprints to advance the field, but the cognitive fingerprint doesn't have a specific definition. We hypothesize that natural pauses (delays between typing characters in words) are caused by cognitive factors (for example, spelling an unfamiliar word or pausing after certain syllables) [5-9], which are unique among individuals. Thus, a cognitive factor can affect the typing rhythm of a specific word.

In this research, each feature is represented by a unique cognitive typing rhythm (CTR), which contains the sequence of digraphs from a specific word. Such features include natural pauses among the CTR's timing information (digraphs, for example) and could be used as a cognitive fingerprint. Conventional keystroke dynamics don't distinguish timing information for different words and only consider a collection of digraphs (such as trigraphs or *n*-graphs). Cognitive factors have been ignored.

Figure 1 shows a collection of digraphs observed for one user. It might seem as if the collection of digraphs represents a part of a keystroke rhythm, but in reality, the digraphs are clustered around different words. For example, we can separate the collection of digraphs "re" according to four different words (*really*, *were*, *parents*, and *store*). This shows that examining digraphs in isolation might result in missing some important information related to specific words. Figure 2 shows two users who both typed the word "really" several times, illustrating the typing rhythm for each.

This observation confirms our hypothesis: a cognitive factor can affect the typing rhythm of a specific word. Thus, we extract CTRs from keystroke dynamics and use them as features (cognitive fingerprints) for active authentication. Each feature is a sequence of digraphs of a specific word (instead of a collection of digraphs). For each legitimate user, we collect samples of each feature and build a classifier for that feature during the machine-learning training phase.

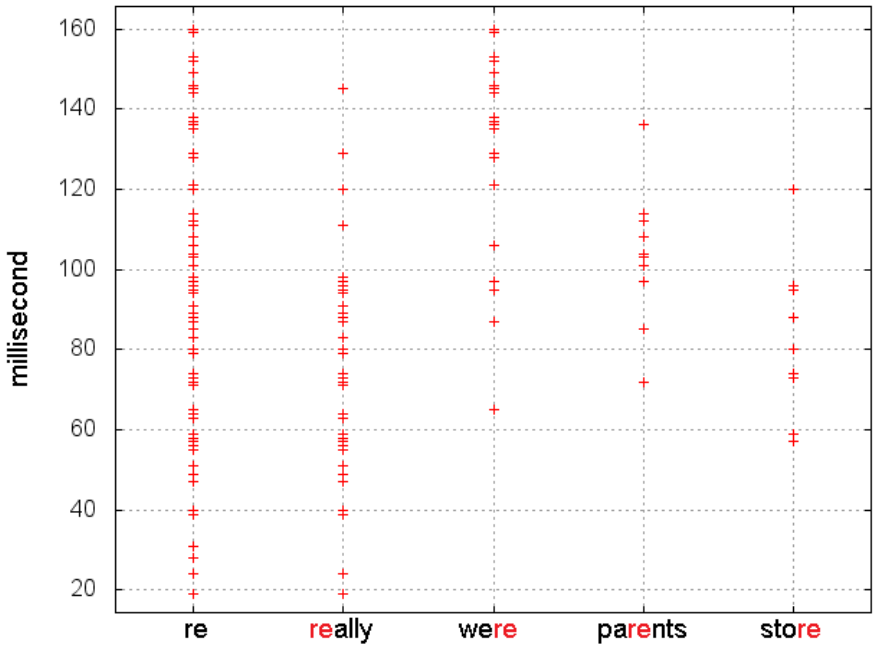

Figure 1. The Digraph "re" from the Same User

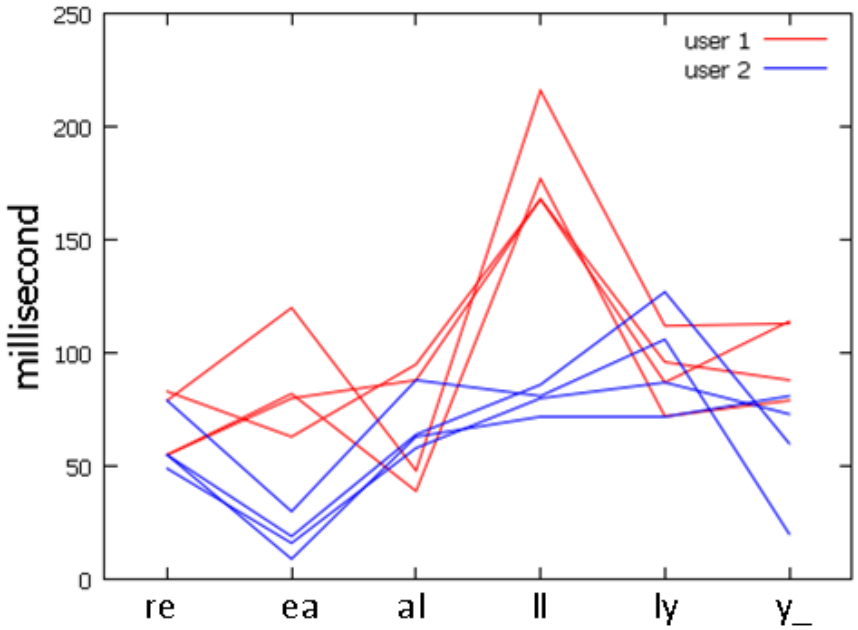

Figure 2. Two Users Typing the Same Word: "really."

3.2 Building an Authentication System

We developed two authentication systems based on two different machine-learning techniques. The first one uses an off-the-shelf support vector machine (SVM) library [10], and the second one employs a library developed in-house, based on kernel ridge regression (KRR) [11]. We used these libraries to build each classifier during the training phase.

Although we can't know the patterns of all imposters, we use patterns from the legitimate user and some known imposters to build each classifier so it can detect a potential imposter. In machine learning, this is known as a two-class (legitimate user vs. imposters) classification approach. We built a trained profile with multiple classifiers for each legitimate user. Then, during the testing phase (authentication), we gave a set of testing data to the trained profile for verification. Each classifier under testing yielded a matching score between the testing dataset and trained file. The final decision (accept or reject) was based on the sum of scores from all classifiers.

The two systems had different basic machine-learning libraries (SVM and KRR) but shared the same feature selection and fusion method. Using the fusion method, we evaluated each classifier to determine the confidence level of its decision. We conducted this evaluation during the training phase using datasets from each legitimate user and from imposters (see Figure 3). We separated the dataset into k equal-sized subsets. Each time, we used $k - 1$ subsets as training data, and we used the remaining subset for testing. We repeated the testing k times, until each subset had been used to test the model. This technique is called k-*fold cross-validation* (or *rotation estimation*).

The test results let us estimate the probabilities of the classifier's true acceptance (P_{ta}) and false acceptance (P_{fa}) rates. For example, after testing with the dataset from a legitimate user, there were N acceptances out of M samples, so P_{ta} is N / M. The confidence of the acceptance decision (W_a) is expressed as the ratio of P_{ta} to P_{fa}. The confidence of the rejection decision (W_r) is expressed as the ratio of the probability of true rejection ($1 - P_{fa}$) to the probability of false rejection ($1 - P_{ta}$).

After the training, in the trained profile, we have W_a and W_r for each classifier. During the testing phase, each classifier generates a decision (acceptance or rejection). Either W_a or W_r will be applied to this decision. The final decision is based on the sum of the scores from all involved classifiers.

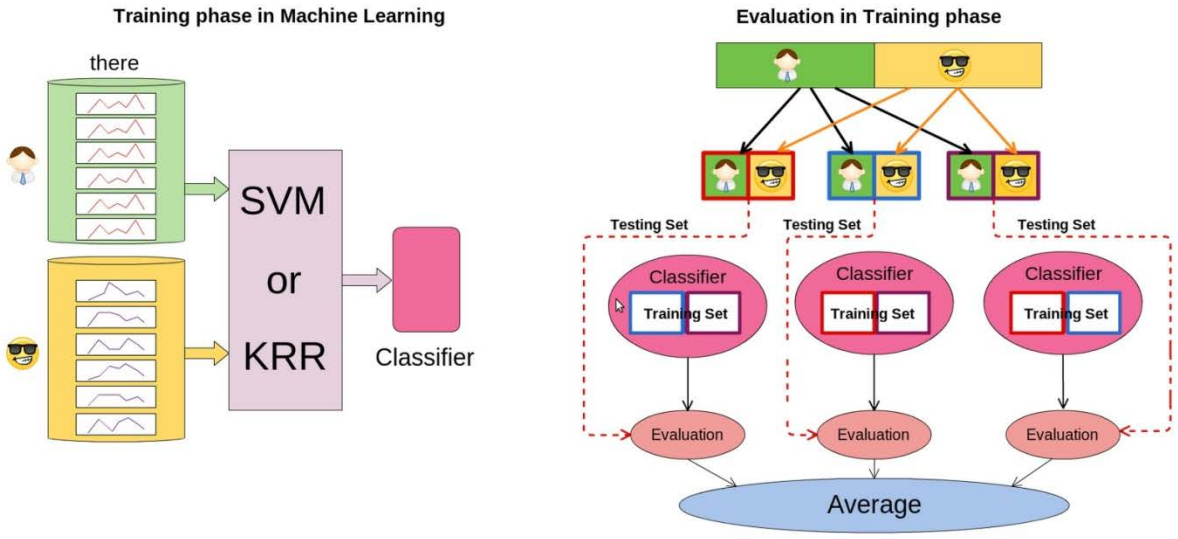

Figure 3. Training and Cross-Validation in Machine Learning: (a) Training Phase for Building a Classifier and (b) Evaluation to Obtain the Confidence of Each Classifier.

3.3 A Large-Scale Experiment

We developed a Web-based software system to collect the keystroke dynamics of individuals in a large-scale testing project conducted at Iowa State University (ISU). This system provided three simulated user environments: typing short sentences, writing short essays, and browsing webpages. We stored the users' cognitive fingerprints in a database for further analyses and applied machine-learning techniques to authenticate users by performing pattern recognition.

During November and December of 2012, we sent email invitations to 36,000 members of the ISU community. There were 1,977 participants who completed two segments, each lasting approximately 30 minutes, resulting in approximately 900 words for each participant for each segment. In addition, 983 participants (out of the 1,977) completed another segment of approximately 30-minutes in length, in which we collected approximately 1,200 words for each participant. We then developed 983 individual profiles (trained files). Each profile was trained under two-class classification, in which one legitimate user had 2,100 collected words, and the imposter training set was based on collected words from the other 982 known participants. Each profile was tested with the data of the 1,977 participants (with a testing dataset of 900 words per participant).

3.4 FURTHER ANALYSIS

At the beginning, we used words as units to extract the biometric features in keystroke dynamics, however, we found some limitations with this approach. The main issue arises when the user never, or seldom, types those exact words in their training profile. The lower possibility that a user uses the same words in the training profile might result in longer testing time. This is mainly because our system would need more data to have a confident result. This would be the case, e.g., if the user typed 100 words but only 20 words can be matched with those words in the training profile. Therefore, we added the use of sub-words to improve our system's performance.

We defined sub-words as the most frequently used typing sequences, and their lengths was from two characters to n (we chose n = 4 in our experiments). For example, if the user types two words: "running" and "walking", and although these are two different words; we consider the features extracted from both words' "ing" as the same feature. With this technique, our system can extract more biometric features with the same amount of data compared to using only words. Hence the system can keep the same accuracy with shorter time of data collection.

4.0 RESULTS AND DISCUSSION

4.1 Initial Results

Table 1 and Figure 4 show the results. Table 1 summarizes the performance comparison of the two verification systems, and Figure 4 shows the detection error trade-off chart from the KRR-based system. In this experiment, each legitimate profile had been tested using the dataset collected from the same user; seven (out of 983) users were recognized as imposters using the SVM library, so it correctly identified the other 976 users, and 17 (out of 983) users with the KRR library, so it correctly identified 966 users. Also, we tested each profile with the other 1,976 participants, and the false-accept rate was 0.055 percent for both SVM and KRR.

Table 1. Performance Comparison of the two Verification Systems.

	SVM	KRR
FAR	0.055	0.055
FRR	0.007	0.0177
training time	15 m/user	29 s/user
testing time	0.6 s/user	3.5 ms/user
size of training file	20 MB/user	1 MB/user

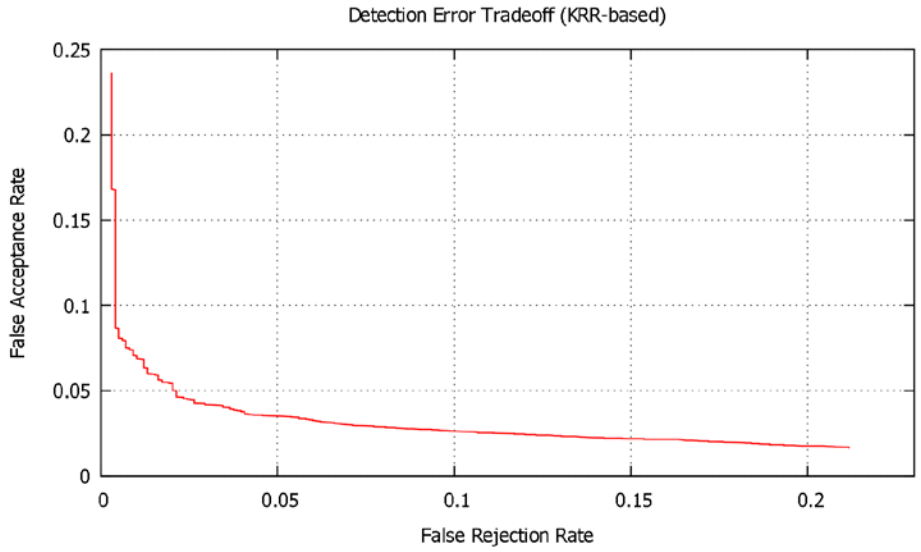

Figure 4. The Detection Error Trade-off Chart from the Kernel-Ridge-Regression-Based System.

4.2 Further Results

Figure 5 shows the comparison with different number of words in testing. Like previous experiment, we used the same number of words for training each profile, but then we used smaller number of words in the testing phase. With this result, we can find the confidential level with different data length (or with different testing window).

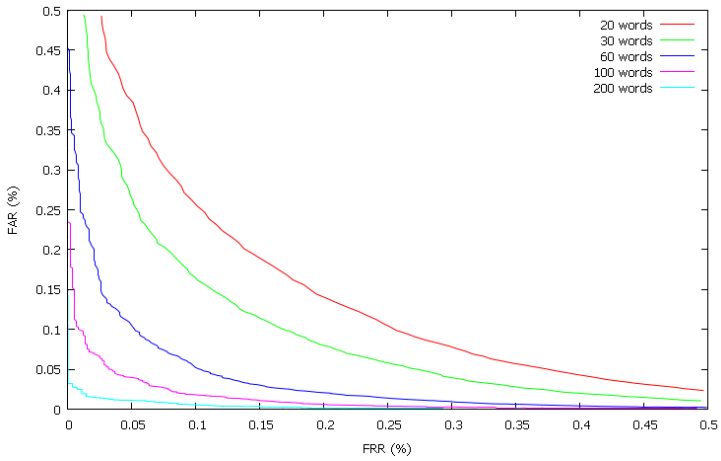

Figure 5. DET Curves per Number of Words

4.3 Live Demo

In our Interactive Cognitive Analysis System, we developed an extension based on Google Chrome browser. At the beginning, the user should login with their ISU Net-ID to our extension. Then, anything that the user types in this browser will be collected and sent to our server. Our sever extracts biometric features from these raw data, and these features will be compared with a specific profile which has been created from previous data collected from the same user. After that, the server will send a feedback to our extension, and show the result on the icon on the upper-right corner of the browser. Green light means that the current user is a legitimate user, red light indicates an imposter, and yellow light means we don't have enough data to make a decision. The live demo was shown first on April 8th, 2013 at the PI meeting in University of Maryland, and was presented to the Active Authentication Community. It was later presented at the Biometric Consortium Conference in Tampa, Florida, on September 17-19, 2013.

Figure 6 shows the extension with the login form, and the red circle is the initial one since the user has not logged in yet. Figure 7 shows the gradual increase and decrease of the score calculation of a certain user based on their typing behavior. The three figures are based on the server calculation of the score based on words, sub-words, and the final fusion between the two. As can be noticed from figure 7, the user appears to be an imposter since their score lies below the zero line. Furthermore, a video of the live demo can be found on the project's website [12].

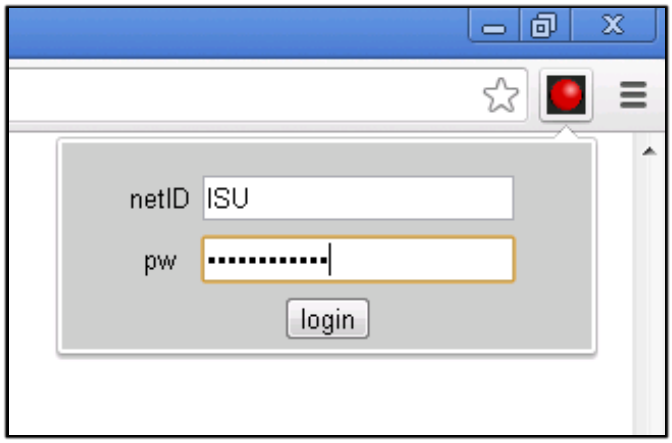

Figure 6. Active Authentication Browser Extension.

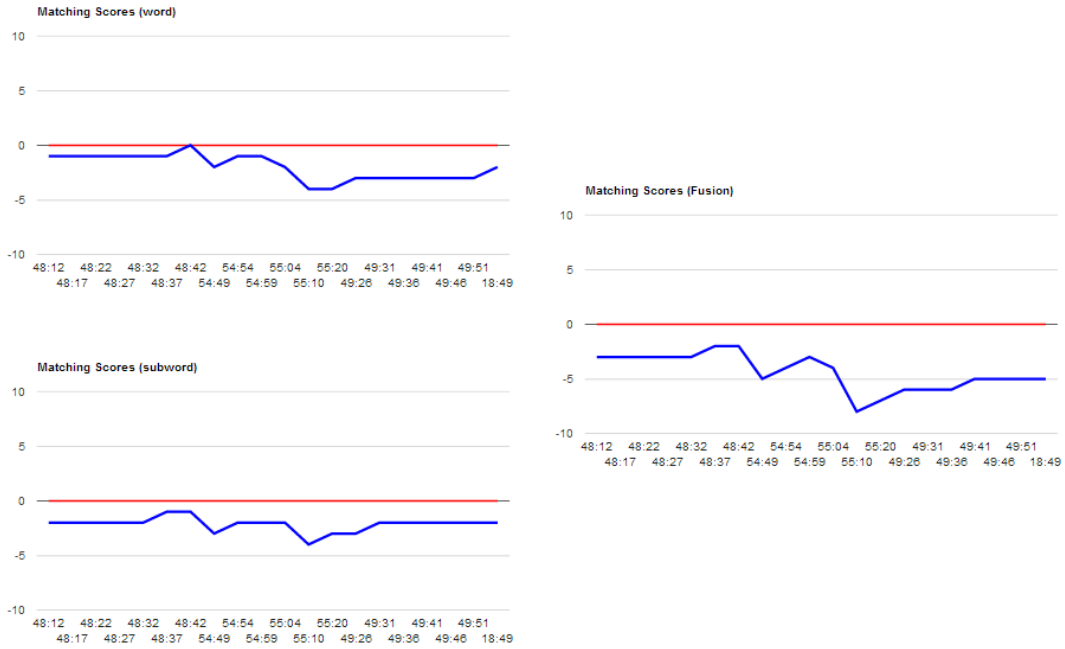

Figure 7. The Demo Webpage Graphs

5.0 CONCLUSIONS

In summary, the proposed scheme is effective for authentication on desktop devices. Moreover, because of the increasing popularity of mobile devices, it's interesting to find the cognitive fingerprint and apply our authentication system on mobile devices. In the future, we'll study keystroke dynamics on different platforms.

Moreover, our live demo was successful at demonstrating the effectiveness of our scheme if it were to be used for active authentication. We intend to add other features and modalities to the live demo in order to improve its accuracy and speed of verifying users.

6.0 REFERENCES

1. Bergadano, F. et al., "User Authentication through Keystroke Dynamics," *ACM Trans. Information and System Security*, Nov. 2002, pp. 367–397.
2. Gunetti, D. and Picardi, C., "Keystroke Analysis of Free Text," *ACM Trans. Information and System Security*, **vol. 8**, no. 3, 2005, pp. 312–347.
3. Karnan, M. et al., "Biometric Personal Authentication Using Keystroke Dynamics: A Review," *Applied Soft Computing*, **vol. 11**, no. 2, 2011, pp. 1565–1573.
4. Monrose, F. et al., "Password Hardening Based on Keystroke Dynamics," *Proc. 6th ACM Conf. Computer and Communications Security*, ACM, 1999, pp. 73–82.
5. Levy, C.M. and Ransdell, S., "Writing Signatures," *The Science of Writing: Theories, Methods, Individual Differences, and Applications*, Lawrence Erlbaum, 1996, pp. 149–162.
6. McCutchen, D., "A Capacity Theory of Writing: Working Memory in Composition," *Educational Psychology Rev.*, **vol. 8**, no. 3, 1996, pp. 299–325.
7. McCutchen, D., "Knowledge, Processing, and Working Memory: Implications for a Theory of Writing," *Educational Psychologist*, **vol. 35**, no. 1, 2000, pp. 13–23.
8. Olive, T., "Working Memory in Writing: Empirical Evidence from the Dual-Task Technique," *European Psychologist*, **vol. 9**, no. 1, 2001, pp. 32–42.
9. Olive, T. et al., "Verbal, Visual, and Spatial Working Memory Demands During Text Composition," *Applied Psycholinguistics*, **vol. 29**, no. 4, 2008, pp. 669–687.
10. Chang, C.-C. and Lin, C.-J, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intelligent Systems and Technology*, **vol. 2**, no. 3, 2011, article no. 27.
11. Kung, S.Y., *Kernel Methods and Machine Learning*, Cambridge Univ. Press, 2013
12. Chang, J.M., "Capturing Cognitive Processing Time for Active Authentication", URL: http://icas.public.iastate.edu/, last modified October 7, 2013.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

CTR	cognitive typing rhythm
DET	detection error tradeoff
FAR	false acceptance rate
FRR	false rejection rate
ISU	Iowa State University
KRR	kernel ridge regression
Net-ID	network identification
P_{fa}	classifier's false acceptance
P_{ta}	classifier's true acceptance
PI	primary investigator
SVM	support vector machine
W_a	acceptance decision
W_r	rejection decision